Supporting Information

Figure S1. Serum miRNA tests.

Methodology. For serum isolation, blood was kept at RT for 30-60 minutes to clot, then spun at 3000 rpm (1000-1300 g) for 10 minutes. The serum was removed and dispensed in 1 ml aliquots into 2 ml cryotubes. Specimens were stored at -80 C. Serum RNA (from 1 ml of sera, 0.5 ml for sera of the symptomatic set) was extracted with Trizol-LS (Invitrogen) combined with mirVana miRNA Isolation Kit (Ambion). Briefly, Trizol-LS was added to serum in volumetric ratios of 3:1 (3 ml Trizol for 1 mL serum) according to the manufacturer's instructions. After the chloroform-addition step and phase separation, the aqueous phase was mixed with 1.25 volumes of absolute ethanol. The resulting solution was then loaded onto the cartridge provided in the mirVana miRNA Isolation Kit (Ambion), and total RNA was eluted in 50 μL of water and stored at -80.

MiRNA levels were analyzed with the TaqMan® Low Density Array microRNA Signature Panel (v1.0), containing 365 human miRNA assays (Applied Biosystems). Briefly, RNA was reverse transcribed using the TaqMan MiRNA Reverse Transcription Kit, and the TaqMan Multiplex RT assays, Human pool set. Each of the eight multiplex RT- reactions was performed using 4.75 μL of extracted RNA (50 μL from 1 mL serum) in a final volume of 20 μL. Each multiplex RT reaction was then diluted to 50 μL with water and combined with 50 μL of TaqMan 2X Universal PCR master Mix, No AmpErase UNG. The final solution (100 μL) was loaded into the array, and RTQ-PCR was carried out on an Applied Biosystems 7900HT thermocycler using the manufacturer's recommended cycling conditions.

A. Selection of miRNA assays. A total of 365 miRNA assays (listed in Table SII of Supporting Information) were initially utilized in the screening. A series of tests was

implemented to exclude those miRNAs that did not meet a number of stringent criteria for inclusion in the final analysis. In step 1 (sequence verification), 10 assays were excluded because they did not exactly match the mature miRNA sequence in the latest miRNA database (miRbase 16.0). In step 2 (profiling feasibility, see also panel B), 161 assays were excluded because they could not be detected reproducibly in sera. In step 3 (quality check 1, see also panels C, E), 32 assays were excluded because they did not show linear behaviour in the detection. In step 4 (quality check 2, see also panels D, F), 15 assays were excluded, because they were possibly affected by preoperative treatments. At the end of these calibrations, 147 miRNA assays were considered suitable for our purposes and subjected to the analyses described in the main text.

B. **Details of Step 2.** We selected only those miRNAs that could be detected reproducibly in sera. For this purpose, we prepared 4 pooled serum samples (2 from normal sera – N1 and N2 pools –, and 2 from tumour sera – T1 and T2 pools –; all sera were from the training set), each containing equal amounts of sera from 5 donors. Total RNA was extracted from each pool and used for miRNA analysis by RTQ-PCR. The data matrix shows the detection of miRNA assays (based on Ct value cut-offs) in the four pools (columns, pools; rows, miRNAs). Assays with a Ct value > 36, were considered not detected, (black), assays with a Ct value < 36 were considered detected (red Ct < 30, orange for Ct 30-36). Only assays detected in at least 2 of 4 pools (194 assays) were retained; the others (161 assays) were excluded from further analyses.

C. **Details of Step 3.** Next, we excluded those miRNA assays that did not display linear behaviour of detection. A pool of sera from 20 healthy individuals was used for this purpose. As an internal control, we added a known amount of a synthetic miRNA (miR-34a, which was never detected in serum) to the pool before dilution (10^7 copies

in 100 μL of total RNA). The known amount of miR-34a was used to calibrate the measurement of miRNA copies in the serum and evaluate the sensitivity of the platform. The pool of sera was then subjected to linear two-fold dilutions, and used for miRNA analysis by RTQ-PCR. Ct values for each miRNA were plotted on a log scale (log quantity against expression Ct) and values were fitted by linear regression. Assays which showed a coefficient of determination (R_2) of < 0.85 were discarded, and 162 assays were retained for further analyses. The data matrix shows the expression profile of miRNA assays in the four dilutions. Examples of individual miRNA assays are shown in panel E.

D. Details of Step 4. The vast majority of sera from the COSMOS study were collected prior to the LD-CT scan. In most cases of NSCLC patients from the COSMOS study, however, tumour sera collected at the baseline were not available; therefore we used sera taken before surgery (which normally took place a few days or weeks after the LD-CT scan). In addition, the sera of the symptomatic set were all taken pre-surgery. In all cases of pre-surgery sera, patients underwent diagnostic procedures before serum collection. Thus, it was important to assess whether a pre-operative bias would affect the levels of some miRNAs; in other words, whether some miRNAs could fluctuate in serum as a result of preoperative diagnostic procedures.

To determine this, we prepared four pools of sera (5 patients/pool) by grouping together matched tumor samples taken from the same LD-CT-detected NSCLC patients at two close time points (<60 days): at the baseline (prior to LD-CT, pool b-T1 and b-T2) and before surgery (pre-operative, pool pre-T1 and pre-T2). Two other pools were generated by pooling sera from healthy individuals, always from COSMOS sera, (pool N1 and N2) and used as a reference to calculate microRNA regulation in each condition. Total RNA was extracted from each pool and used for

microRNA analysis by RTQ-PCR. The data matrix shows the regulation of each microRNA at the baseline (pool b-T over pool N) and at preoperative stage (pool pre-T over pool N) in two different experiments (EXP1, N1 vs. b-T1 vs. pre-T1; EXP2, N2 vs. b-T2 vs. pre-T2). A total of 15 miRNA assays were reproducibly regulated only at the preoperative stage and were, therefore, excluded from further analyses. Examples are shown in panel F. **E**. examples of miRNA selection at step 3. **F**. Examples of miRNA selection at step 4.

Figure S2. **Serum miRNA dataset**. Expression profile data matrix of the 147 miRNAs in all the analyzed sera. A total of 253 sera (whose origin is indicated in the legend) were profiled. Data matrices show the Ct values of each microRNA before (A) and after (B) normalization. Coloured bars indicate the class of each patient. Columns, patients; rows, miRNAs.

Figure S3. **Cluster analysis**. One-way hierarchical clustering of the expression profile of the 34 miRNAs in the training and testing sets. Columns represent log2 ratios of expression of each miRNA; rows represent patients. Coloured bars indicate the class of each patient. AC, adenocarcinoma; SCC, squamous cell carcinoma. Two classes of miRNA, showing decreased (cluster A) or increased (cluster B) expression in tumour vs. normal sera were consistently detected in training and testing sets.

Figure S4. **Comparison between different normalization procedures.** To further control for the robustness of our findings, we repeated the evaluation of the performance of the 34-miRNA model using a different method for normalization, the "median" method in which data were centred based on the median Ct value of all the

147 detected miRNAs. The figure shows the various risk indexes, based on the 34-miRNA model, with data normalized by the housekeeping method (HK normalized, on the left) compared with data normalized by median Ct (Median normalized, on the right), for samples analyzed and shown in Figs 2 of the main text. A correlation analysis between Risk indexes of HK-normalized and Median-normalized data is also shown.

Figure S5. **Performances of predictor models using different combinations of miRNAs.**

The performance (in terms of AUC) of various predictor models (from 5 to 34 miRNAs) in the testing set is shown. MiRNAs were selected within the 34 miRNA list before the training of the classifier, and performances were measured by DLDA (diagonal linear discriminant analysis) and K-fold (K=5) cross-validation method repeated 100 times. MiRNA composition of the various predictor models is shown in the lower panel.

Table SI. Clinical and pathological information for all patients and gene weights and data used for prediction rule.

[Table SI is provided as a separate .xls file]

Legend to Table SI. For each patient (ID 1 through 253), the following information is provided: Class (BDO, breast nodule, lung hamartoma, normal, lung nodule, lung tumour); BDP/Tumour match (showing those sera representing matching pairs before and after the disease onset); Set (BDO, breast, PH, training, testing set, symptomatic set, nodules); Gender (M/F); Age (years); Type (histological type, AC or SCC, for NSCLC only); Stage (I through IV, for NSCLC only); Nodule type (for lung nodules only); Smoking status; Pack/year (number of packs of cigarettes smoked per day multiplied by the number of years the person has smoked); Risk, 34-miRNA model derived risks. Wi, gene weights.

Table SII. Synopsis of all experiments and characterizations performed with the 365 miRNA assays.

[Table SII is provided as a separate .xls file]

Legend to Table SII. The table shows data on the 365 miRNA assays used in the serum miRNA profile. Shown are the assay ID, the accession number, the ID in the miRbase 16 database, the mature miRNA sequence, and the results of tests described in Figure S1 of Supporting Information. Also shown are the miRNAs that were selected as "housekeeping" for normalization (HK) and those selected as markers by DLDA. The 147 miRNA assays retained in all presented analysis are in red.

Table SIII. Serum housekeeping miRNAs used for the normalization procedures

ASSAY ID	Accession	Mature_Seq	miRbase 16	E (R^2)	CT	%	P	SD
hsa-miR-146a-4373132	MI0000477	UGAGAACUGAAUUCCAUGGGUU	hsa-mir-146a	1.00	28.57	100	0.44	0.82
hsa-miR-15b-4373122	MI0000438	UAGCAGCACAUCAUGGUUUACA	hsa-mir-15b	1.00	28.98	100	0.42	0.88
hsa-miR-197-4373102	MI0000239	UUCACCACCUUCUCCACCCAGC	hsa-mir-197	0.99	25.98	100	0.14	0.74
hsa-miR-19a-4373099	MI0000073	UGUGCAAAUCUAUGCAAAACUGA	hsa-mir-19a	1.00	29.97	100	0.69	0.78
hsa-miR-19b-4373098	MI0000074	UGUGCAAAUCCAUGCAAAACUGA	hsa-mir-19b	1.00	26.17	100	0.45	0.72
hsa-miR-24-4373072	MI0000080	UGGCUCAGUUCAGCAGGAACAG	hsa-mir-24	1.00	26.79	100	0.90	0.86

Legend to Table SIII. The serum housekeeping miRNAs selected in the training set are shown (with indication of the assay ID, accession number, sequence and ID in miRbase16). E, efficiency of the RTQ assay, expressed as coefficient of determination (R^2), and determined as illustrated in Figure S1 of Supporting Information. Also shown are the parameters used for the selection of the HK in the training cohort: Ct (median) of expression (which had to be less than 30 for inclusion); percentage (%) of sample with RTQ detection (which had to be 100% for inclusion), p-value (P) by Welch T-test in the training set (which had to be > 0.1 for inclusion), and Ct standard deviation (SD) in the training set (which had to be < 0.9 for inclusion).

Table SIV. The 34-miRNA model

Assay	Accession	Sequence	miRbase 16	Fold	P-value
hsa-miR-92-4373013	MI0000094	UAUUGCACUUGUCCCGGCCUGU	hsa-mir-92a	-1.34	0.04
hsa-miR-484-4381032	MI0002468	UCAGGCUCAGUCCCCUCCCGAU	hsa-mir-484	-1.25	0.01
hsa-miR-486-4378096	MI0002470	UCCUGUACUGAGCUGCCCCGAG	hsa-mir-486-5p	-1.67	0.001
hsa-miR-328-4373049	MI0000804	CUGGCCCUCUCUGCCCUUCCGU	hsa-mir-328	-1.33	0.04
hsa-miR-191-4373109	MI0000465	CAACGGAAUCCCAAAAGCAGCUG	hsa-mir-191	-1.19	0.02
hsa-miR-376a-4373026	MI0003529	AUCAUAGAGGAAAAUCCACGU	hsa-mir-376a	-4.03	0.02
hsa-miR-342-4373040	MI0000805	UCUCACACAGAAAUCGCACCCGU	hsa-mir-342-3p	-1.60	<0.001
hsa-miR-331-4373046	MI0000812	GCCCCUGGGCCUAUCCUAGAA	hsa-mir-331-3p	-1.36	<0.001
hsa-miR-30c-4373060	MI0000736	UGUAAACAUCCUACACUCUCAGC	hsa-mir-30c	-1.73	0.006
hsa-miR-28-4373067	MI0000086	AAGGAGCUCACAGUCUAUUGAG	hsa-mir-28-5p	-1.67	0.02
hsa-miR-98-4373009	MI0000100	UGAGGUAGUAAGUUGUAUUGUU	hsa-mir-98	-1.73	0.04
hsa-miR-17-5p-4373119	MI0000071	CAAAGUGCUUACAGUGCAGGUAG	hsa-mir-17	-1.63	<0.001
hsa-miR-26b-4373069	MI0000084	UUCAAGUAAUUCAGGAUAGGU	hsa-mir-26b	-1.44	0.004
hsa-miR-374-4373028	MI0000782	UUAUAAUACAACCUGAUAAGUG	hsa-miR-374a	-1.55	0.02
hsa-miR-30b-4373290	MI0000441	UGUAAACAUCCUACACUCAGCU	hsa-mir-30b	-1.56	0.002
hsa-miR-26a-4373070	MI0000083	UUCAAGUAAUCCAGGAUAGGCU	hsa-mir-26a	-1.36	0.02
hsa-miR-142-3p-4373136	MI0000458	UGUAGUGUUUCCUACUUUAUGGA	hsa-mir-142-3p	-1.62	0.009
hsa-miR-103-4373158	MI0000108	AGCAGCAUUGUACAGGGCUAUGA	hsa-mir-103	-1.71	<0.001
hsa-miR-126-4378064	MI0000471	UCGUACCGUGAGUAAUAAUGCG	hsa-mir-126	-1.25	0.02
hsa-let-7a-4373169	MI0000060	UGAGGUAGUAGGUUGUAUAGUU	hsa-let-7a	-1.87	<0.001
hsa-let-7d-4373166	MI0000065	AGAGGUAGUAGGUUGCAUAGUU	hsa-let-7d	-1.48	0.05
hsa-let-7b-4373168	MI0000063	UGAGGUAGUAGGUUGUGUGGUU	hsa-let-7b	-1.64	<0.001
hsa-miR-32-4373056	MI0000090	UAUUGCACAUUACUAAGUUGCA	hsa-mir-32	2.19	0.007
hsa-miR-133b-4373172	MI0000822	UUUGGUCCCCUUCAACCAGCUA	hsa-mir-133b	2.36	0.04
hsa-miR-566-4380943	MI0003572	GGGCGCCUGUGAUCCCAAC	hsa-mir-566	2.95	0.02

hsa-miR-432-4378076	MI0003133	CUGGAUGGCUCCUCCAUGUCU	hsa-mir-432*	2.66	<0.001
hsa-miR-223-4373075	MI0000300	UGUCAGUUUGUCAAAUACCCCA	hsa-mir-223	1.94	<0.001
hsa-miR-29a-4373065	MI0000087	UAGCACCAUCUGAAAUCGGUUA	hsa-mir-29a	1.66	0.01
hsa-miR-148a-4373130	MI0000253	UCAGUGCACUACAGAACUUUGU	hsa-mir-148a	1.84	0.02
hsa-miR-142-5p-4373135	MI0000458	CAUAAAGUAGAAAGCACUACU	hsa-mir-142-5p	1.34	0.04
hsa-miR-22-4373079	MI0000078	AAGCUGCCAGUUGAAGAACUGU	hsa-mir-22	-1.67	0.02
hsa-miR-148b-4373129	MI0000811	UCAGUGCAUCACAGAACUUUGU	hsa-mir-148b	-1.59	0.04
hsa-miR-140-4373138	MI0000456	CAGUGGUUUUACCCUAUGGUAG	hsa-mir-140-5p	1.49	<0.001
hsa-miR-139-4373176	MI0000261	UCUACAGUGCACGUGUCUCCAG	hsa-mir-139-5p	-2.36	<0.001

Legend to Table SIV. The composition of the 34-miRNA model includes the identity of the relative miRNA assay, accession number, sequence and nomenclature in miRbase 16. Fold change and p-value (parametric t-test) are relative to the expression of miRNAs in the 25 tumour sera versus the 39 normal sera of the training set.

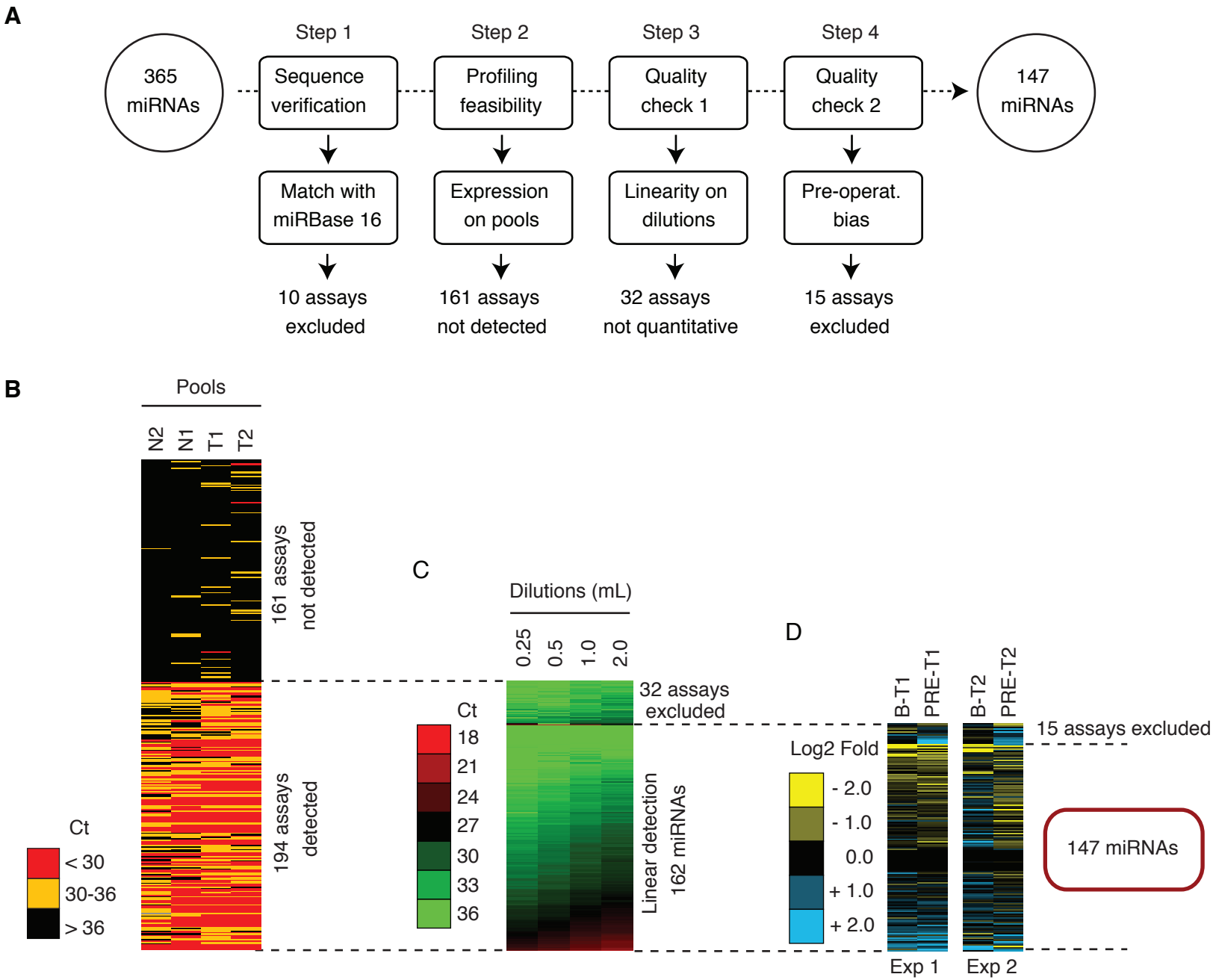
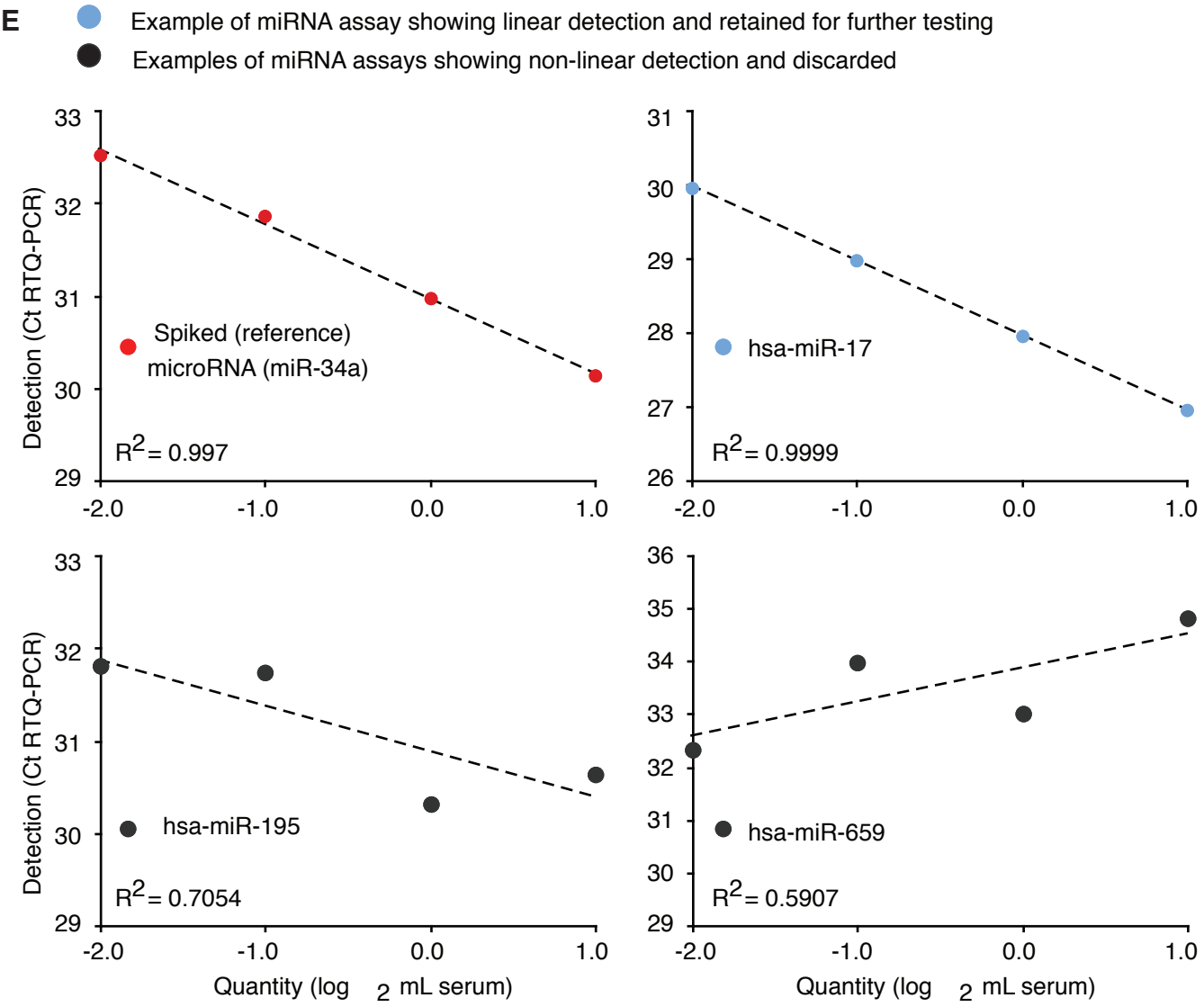
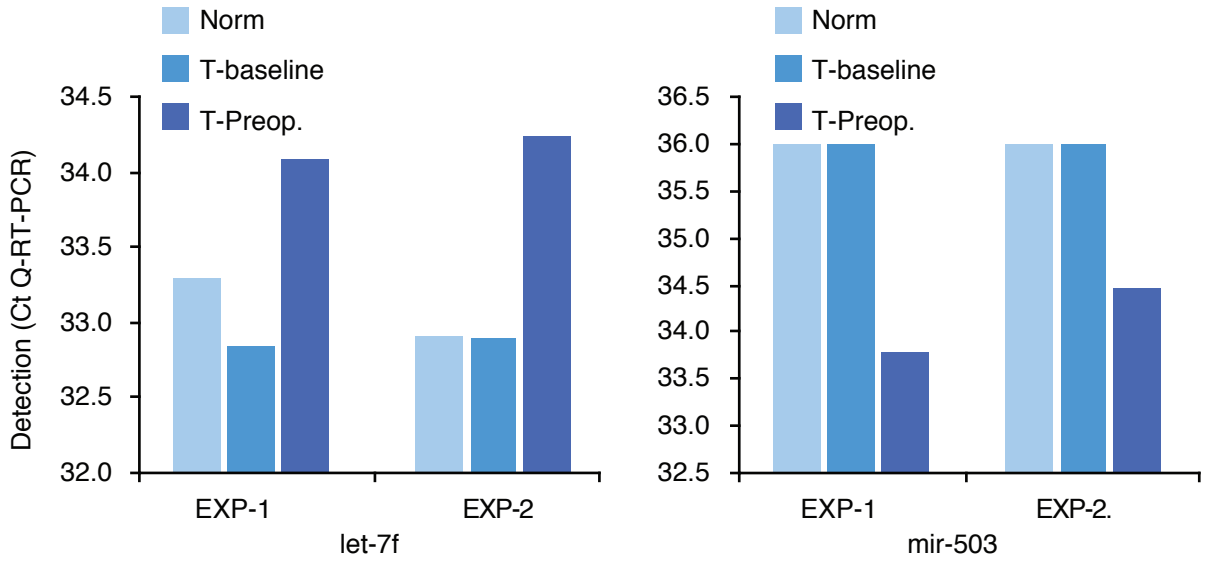

A
365 miRNAs
Step 1
Sequence verification
Match with miRBase 16
10 assays excluded
Step 2
Profiling feasibility
Expression on pools
161 assays not detected
Step 3
Quality check 1
Linearity on dilutions
32 assays not quantitative
Step 4
Quality check 2
Pre-operat. bias
15 assays excluded
147 miRNAs
B
Pools
N2
N1
T1
T2
161 assays not detected
194 assays detected
Ct
< 30
30-36
> 36
C
Dilutions (mL)
0.25
0.5
1.0
2.0
Ct
18
21
24
27
30
33
36
32 assays excluded
Linear detection 162 miRNAs
D
B-T1
PRE-T1
B-T2
PRE-T2
Log2 Fold
- 2.0
- 1.0
0.0
+ 1.0
+ 2.0
15 assays excluded
147 miRNAs
Exp 1
Exp 2

Figure S1A-D

E



F



Examples of miRNA assays showing pre-operative bias and discarded

Figure S1E-F

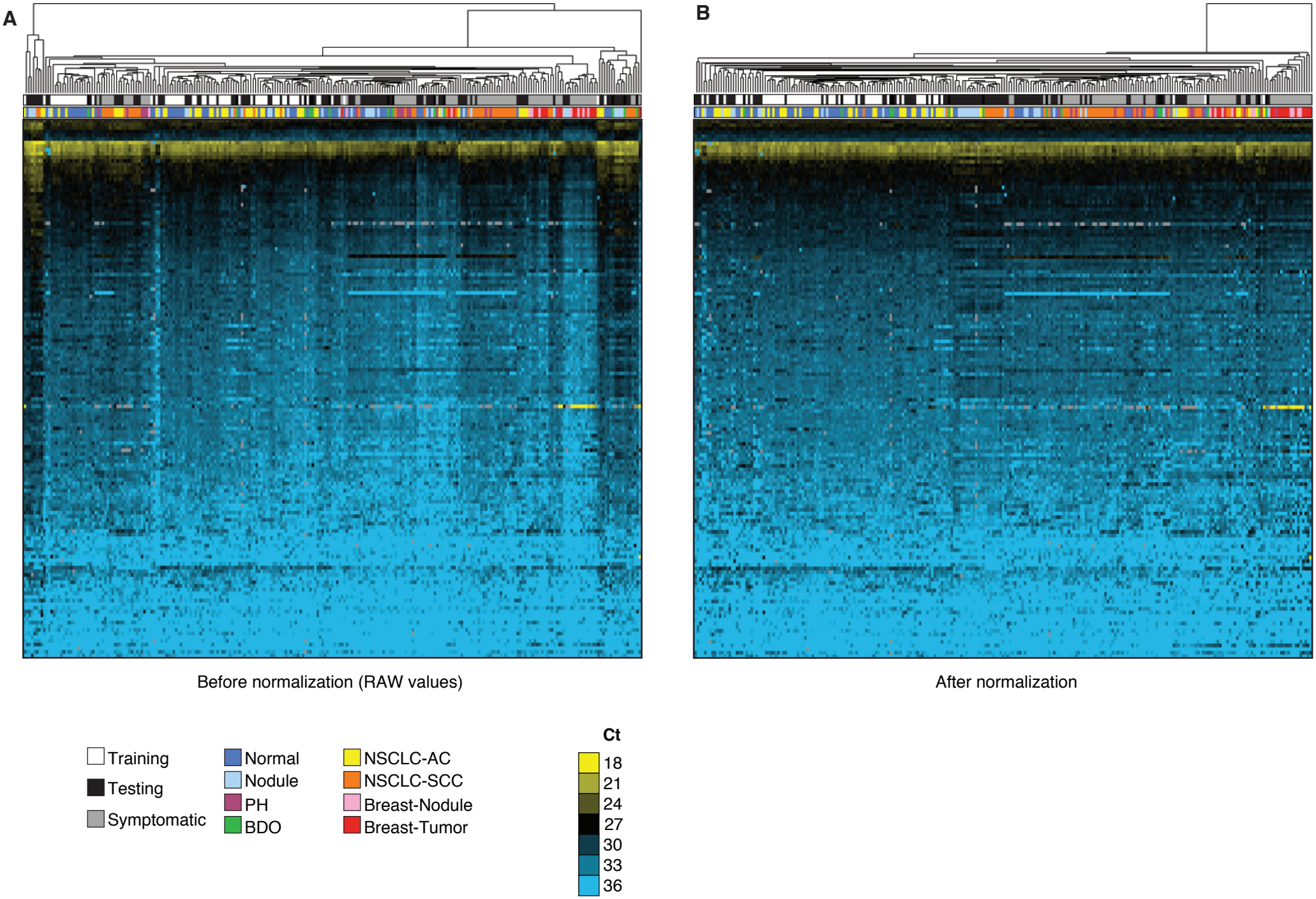

A
B
Before normalization (RAW values)
After normalization
Training
Testing
Symptomatic
Normal
Nodule
PH
BDO
NSCLC-AC
NSCLC-SCC
Breast-Nodule
Breast-Tumor
Ct
18
21
24
27
30
33
36

Figure S2

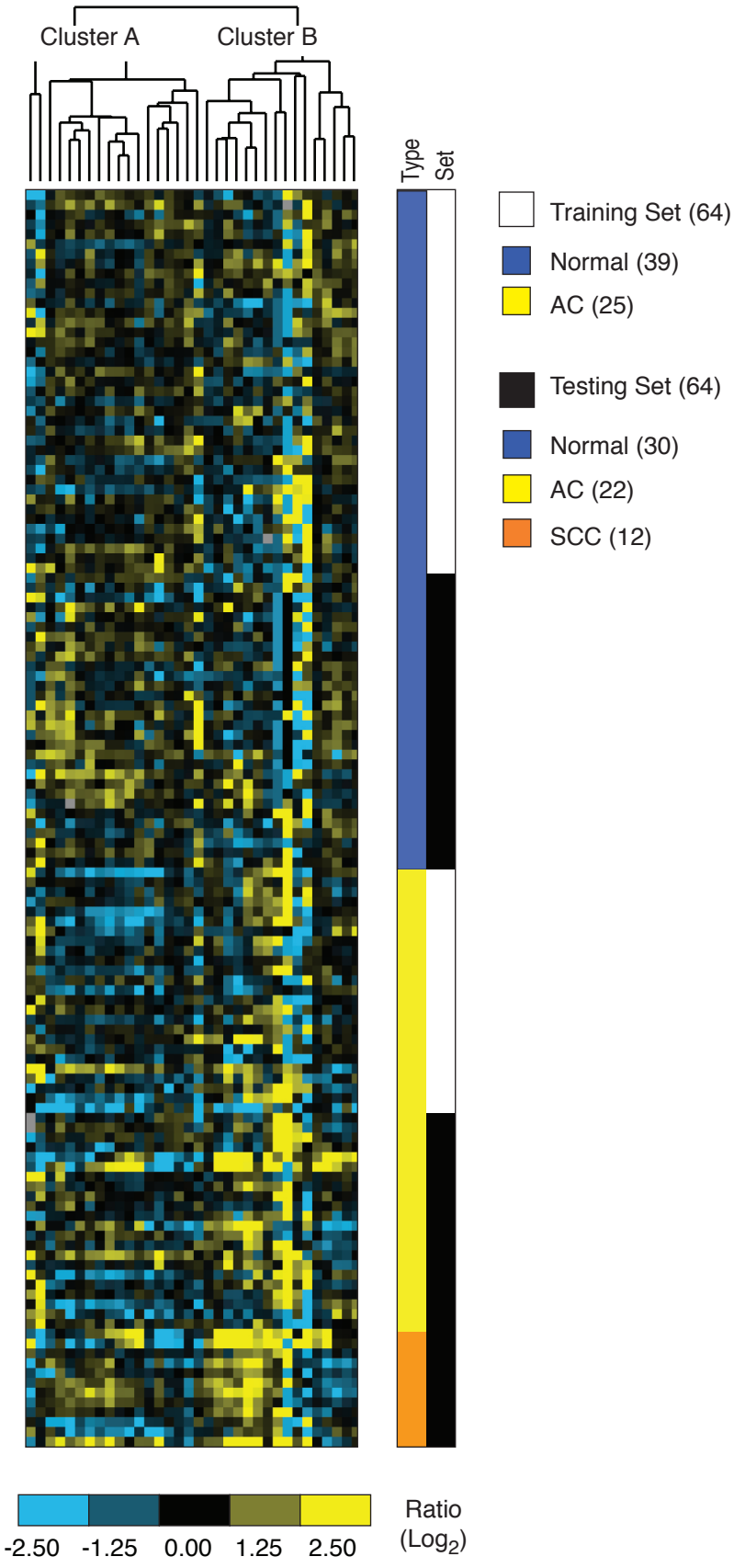
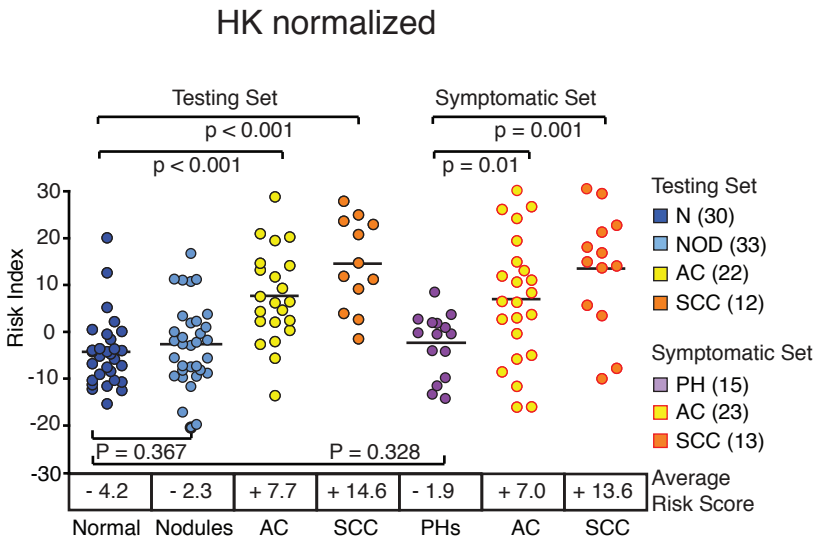
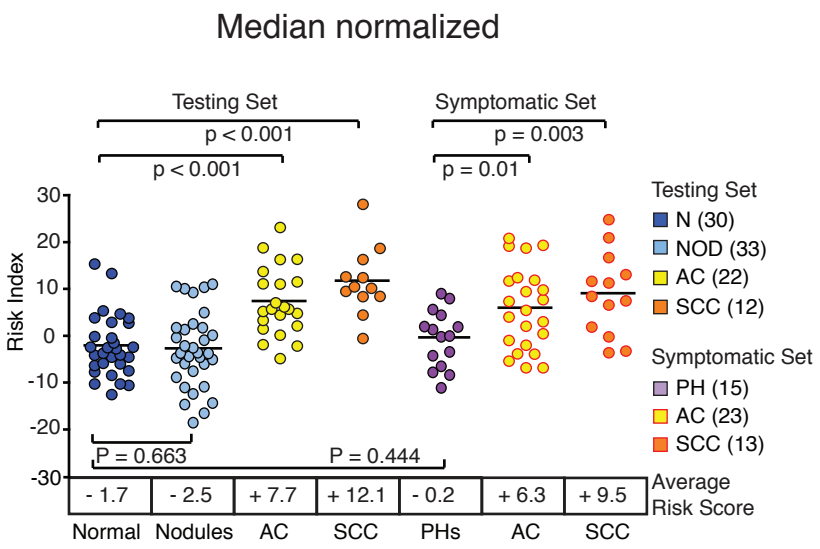

Cluster A
Cluster B
Type
Set
Training Set (64)
Normal (39)
AC (25)
Testing Set (64)
Normal (30)
AC (22)
SCC (12)
-2.50
-1.25
0.00
1.25
2.50
Ratio
(Log2)

Figure S3


HK normalized
Testing Set
Symptomatic Set
p < 0.001
p < 0.001
p = 0.001
p = 0.01
Risk Index
30
20
10
0
-10
-20
-30
P = 0.367
P = 0.328
Testing Set
N (30)
NOD (33)
AC (22)
SCC (12)
Symptomatic Set
PH (15)
AC (23)
SCC (13)
- 4.2
- 2.3
+ 7.7
+ 14.6
- 1.9
+ 7.0
+ 13.6
Average Risk Score
Normal
Nodules
AC
SCC
PHs
AC
SCC


Median normalized
Testing Set
Symptomatic Set
p < 0.001
p < 0.001
p = 0.003
p = 0.01
Risk Index
30
20
10
0
-10
-20
-30
P = 0.663
P = 0.444
Testing Set
N (30)
NOD (33)
AC (22)
SCC (12)
Symptomatic Set
PH (15)
AC (23)
SCC (13)
- 1.7
- 2.5
+ 7.7
+ 12.1
- 0.2
+ 6.3
+ 9.5
Average Risk Score
Normal
Nodules
AC
SCC
PHs
AC
SCC

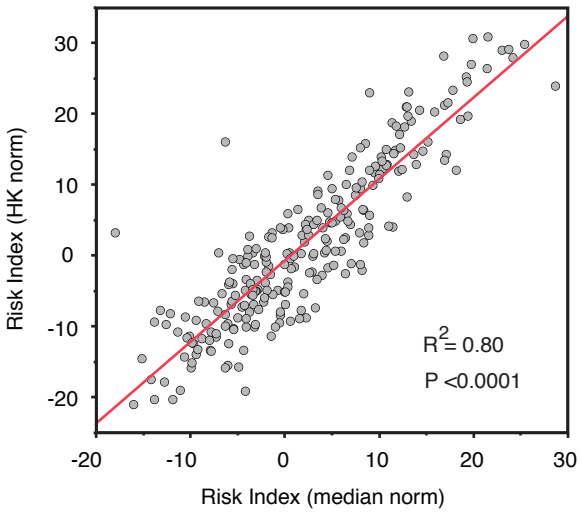
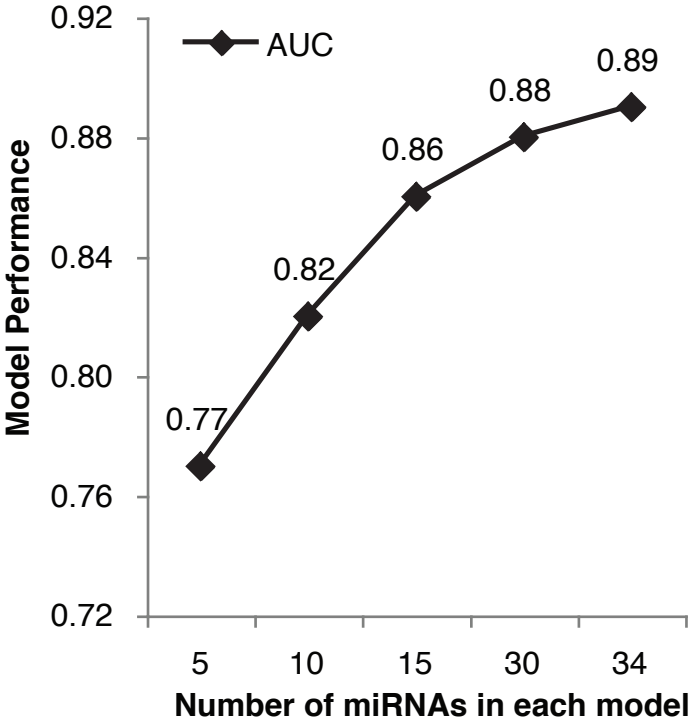

Risk Index (HK norm)
30
20
10
0
-10
-20
-20
-10
0
10
20
30
Risk Index (median norm)
R2= 0.80
P <0.0001

Figure S4


0.92
0.88
0.84
0.80
0.76
0.72
AUC
0.89
0.88
0.86
0.82
0.77
Model Performance
5
10
15
30
34
Number of miRNAs in each model

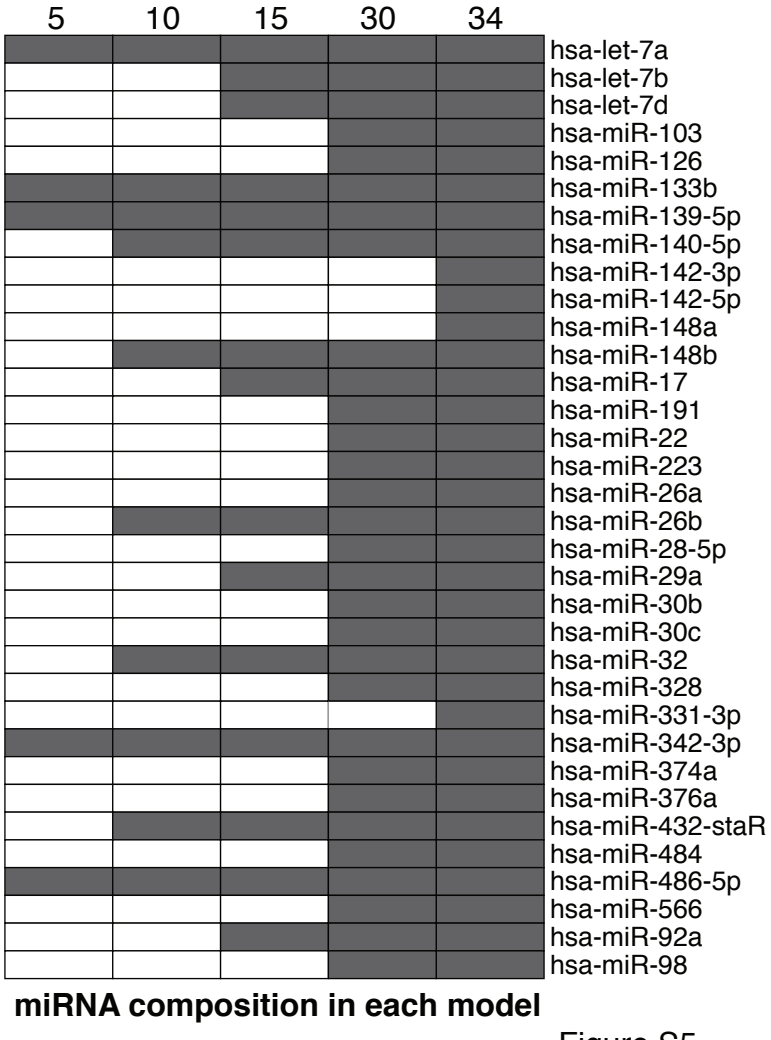

5
10
15
30
34
hsa-let-7a
hsa-let-7b
hsa-let-7d
hsa-miR-103
hsa-miR-126
hsa-miR-133b
hsa-miR-139-5p
hsa-miR-140-5p
hsa-miR-142-3p
hsa-miR-142-5p
hsa-miR-148a
hsa-miR-148b
hsa-miR-17
hsa-miR-191
hsa-miR-22
hsa-miR-223
hsa-miR-26a
hsa-miR-26b
hsa-miR-28-5p
hsa-miR-29a
hsa-miR-30b
hsa-miR-30c
hsa-miR-32
hsa-miR-328
hsa-miR-331-3p
hsa-miR-342-3p
hsa-miR-374a
hsa-miR-376a
hsa-miR-432-staR
hsa-miR-484
hsa-miR-486-5p
hsa-miR-566
hsa-miR-92a
hsa-miR-98
miRNA composition in each model

Figure S5